

WHAT IS CLAIMED IS:

1. A method for selecting a combination of nucleic acid sample pairs for evaluating the ability of an oligonucleotide probe to measure differential expression of genes, said method comprising:

5 (a) conducting differential expression experiments using (i) nucleic acid sample pairs and (ii) nucleic acid probes immobilized on a substrate, said probes representing a set of genes where the number of genes in the set is a portion of an expected number of genes in a sample, and

10 (b) selecting a nucleic acid sample pair combination in relation to the members of said combination having a maximized number of genes from the set of genes that exhibit differential expression and a minimized number of said genes that do not exhibit differential expression.

15 2. A method according to Claim 1 wherein a determination is made for each gene in step (a) whether the gene is differentially expressed and whether probes for each gene cluster together.

20 3. A method according to Claim 2 wherein said determination is made based on one or more parameters from said differential expression experiments.

4. A method according to Claim 3 wherein said parameters are selected from the group consisting of LogRatio, LogRatio error and signal intensities.

25 5. A method according to Claim 3 wherein said parameters are (i) the probability of a combined LogRatio value being significantly different from zero for each of said probes representing a gene and (i) the number of probes for a gene that have a probability of the combined LogRatio value being significantly different from zero above a threshold value.

30 6. A method according to Claim 1 wherein said nucleic acid sample pairs are tissue pairs.

7. A method according to Claim 1 wherein in step (a) said differential expression experiments are conducted by contacting a nucleic acid sample pair with a substrate having said nucleic acid probes immobilized thereon.

5 8. A method according to Claim 1 wherein in step (a) said differential expression experiments are conducted by contacting each member of a nucleic acid sample pair with a separate substrate having said nucleic acid probes immobilized thereon.

10 9. A method for selecting a combination of nucleic acid sample pairs for evaluating the ability of an oligonucleotide probe to measure differential expression of genes, said method comprising:

15 (a) contacting each nucleic acid sample pair from a plurality of nucleic acid sample pairs with a plurality of probes for each of a predetermined number of genes to determine whether said genes exhibit differential expression,

(b) determining for each gene and each of said nucleic acid sample pairs whether said gene exhibits or does not exhibit differential expression based on one or more parameters,

20 (c) for a gene that exhibits differential expression in step (b) for a nucleic acid sample pair, assigning a "yes" value, and for each gene that does not exhibit differential expression in step (b) for a nucleic acid sample pair, assigning a "no" value,

(d) tabulating data from step (c) for each combination of said nucleic acid sample pairs to be evaluated, and

25 (e) selecting a combination of nucleic acid sample pairs having a score based on a maximized number of "yes's" and a minimized number of "no's."

10. A method according to Claim 9 wherein said parameters are selected from the group consisting of LogRatio, LogRatio error and signal intensities.

30 11. A method according to Claim 9 wherein said nucleic acid sample pairs are tissue pairs.

12. A computer-readable medium having recorded thereon a program that

carries out steps (b)-(e) of a method according to the method of Claim 9.

13. A computational analysis system comprising a computer-readable medium according to Claim 12.

5

14. A method for selecting a combination of nucleic acid sample pairs for evaluating the ability of an oligonucleotide probe to measure differential expression of genes, said method comprising:

10 (a) contacting each nucleic acid sample pair from a plurality of nucleic acid sample pairs with a plurality of probes for each of a predetermined number of genes to determine whether said genes exhibit differential expression,

(b) determining for each gene and each of said nucleic acid sample pairs whether said gene exhibits or does not exhibit differential expression based on one or more parameters,

15 (c) analyzing the results of said contacting and said determining utilizing an analysis system comprising a computer, said computer comprising:

(i) means for assigning, for each gene and to each of said nucleic acid sample pairs, a value of 1 if said gene exhibits differential expression or assigning a value of 0 if said gene does not exhibit differential expression,

20 (ii) means for tabulating data from step (i) to determine the total number scores for each combination of said nucleic acid sample pairs to be evaluated, wherein the maximum number of 1's for each nucleic acid sample pair per gene is the number "n" of nucleic acid sample pairs of said combination, and

25 (iii) means for selecting a combination of nucleic acid sample pairs based on a score representing a maximized number of n's and a minimized number of 0's.

15. A method according to Claim 14 wherein said parameters are selected from the group consisting of LogRatio, LogRatio error and signal intensities.

30

16. A method according to Claim 14 wherein said nucleic acid sample pairs are tissue pairs.

17. A computer-based method for selecting a combination of nucleic acid sample pairs for evaluating the ability of an oligonucleotide probe to measure differential expression of genes, said method comprising:

(a) contacting each nucleic acid sample pair from a plurality of nucleic acid sample pairs with a plurality of probes for each of a predetermined number of genes to determine whether said genes exhibit differential expression,

(b) under computer control, determining for each gene and each of said nucleic acid sample pairs whether said gene exhibits or does not exhibit differential expression based on one or more parameters,

(c) under computer control, for a gene that exhibits differential expression in step (b) for a nucleic acid sample pair, assigning a "yes" value, and for each gene that does not exhibit differential expression in step (b) for a nucleic acid sample pair, assigning a "no" value,

(d) under computer control, tabulating data from step (c) for each combination of said nucleic acid sample pairs to be evaluated, and

(f) under computer control, selecting a combination of nucleic acid sample pairs having a score based on a maximized number of "yes's" and a minimized number of "no's."

18. A method according to Claim 17 wherein said parameters are selected from the group consisting of LogRatio, LogRatio error and signal intensities.

19. A method according to Claim 17 wherein said nucleic acid sample pairs are tissue pairs.

20. A method of identifying a sequence of a nucleic acid that is suitable for use as a substrate surface immobilized probe for a target nucleic acid, said method comprising evaluating the sequence using nucleic acid sample pairs selected by a method of claim 1.

21. A method of identifying a sequence of a nucleic acid that is suitable for use as a substrate surface immobilized probe for a target nucleic acid, said method comprising:

- (a) identifying a plurality of candidate probe sequences for said target nucleic acid based on at least one selection criterion;
- (b) empirically evaluating each of said candidate probe sequences under a plurality of different experimental sets to obtain a collection of empirical data values for each of said candidate nucleic acid probe sequences for each of said plurality of different experimental sets wherein said empirical evaluation employs a nucleic acid sample pair selected by a method according to Claim 1;
- (c) clustering said candidate probe sequences into one or more groups of candidate probe sequences based on each candidate probe sequence's collection of empirical data values, wherein each of said one or more groups exhibits substantially the same performance across said plurality of experimental sets;
- (d) selecting one of said one or more groups based on at least one criterion; and
- (e) choosing a candidate probe sequence from said selected group to as said sequence of said nucleic acid that is suitable for use as a substrate immobilized probe for said target nucleic acid.

22. A method of producing an array of nucleic acids on the surface of a substrate, said method comprising:

- (a) identifying nucleic acid probes by a method according to Claim 21 and
- (b) synthesizing or depositing said nucleic acid probes identified in step (a) in an array on the surface of a substrate.

23. A method of detecting the presence of a nucleic acid analyte in a sample, said method comprising:

- (a) contacting a nucleic acid array produced according to Claim 22 with said sample and
- (b) detecting the presence of binding complexes on the surface of said array to detect the presence of said analyte in said sample.

30 24. A method comprising forwarding data representing a result obtained from a method of Claim 23.

25. A method according to claim 23 wherein the data is transmitted to a remote location.

5 26. A method comprising receiving data representing a result obtained from a method of Claim 23.

27. A kit for selecting a combination of nucleic acid sample pairs for evaluating the ability of an oligonucleotide probe to measure differential expression of genes, said kit comprising in packaged combination:

10 (a) an algorithm for use in conducting differential gene expression experiments using (i) nucleic acid sample pairs and (ii) nucleic acid probes immobilized on a substrate where the probes represent a set of genes, the number of genes in the set being a portion of an expected number of genes in a sample, said algorithm being present on a computer readable medium, and

15 (b) instructions for using the algorithm to select a nucleic acid sample pair combination in relation to the members of said combination having a maximized number of genes from the set of genes that exhibit differential expression and a minimized number of said genes that do not exhibit differential expression.

20 28. A kit for identifying a sequence of a nucleic acid that is suitable for use as a substrate surface immobilized probe for a target nucleic acid, said kit comprising:

25 (a) an algorithm that identifies a sequence of a nucleic acid that is suitable for use as a substrate surface immobilized probe for said target nucleic acid according to the method according to Claim 21, wherein said algorithm is present on a computer readable medium; and

(b) instructions for using said algorithm to identify said sequence of a nucleic acid that is suitable for use as a substrate surface immobilized probe for said target nucleic acid.